



**UN ESEMPIO DI ANALISI DEI TESTI: TORINO
CITTA' DELLE ALPI**

Gianluca Bo

Relazione al convegno

*Data Mining e Metodologia della Ricerca Sociale: la Creazione di Valore
Aggiunto per l'Utente, Torino, Aula Magna del Rettorato
dell'Università, 20 Febbraio 2004*

**DATA MINING E METODOLOGIA DELLA RICERCA SOCIALE:
LA CREAZIONE DI VALORE AGGIUNTO PER L'UTENTE**

Negli ultimi anni si è assistito ad un crescente interesse da parte della Business Intelligence alle tecniche di analisi dei testi.

La crescente quantità di informazioni di tipo testuale facilmente disponibili per le organizzazioni (si pensi ad esempio allo sviluppo di Internet) ha di fatto reso prezioso l'affiancamento di tecniche di text mining ai più tradizionali strumenti di analisi dati.

Così come le tecniche di Data Mining consentono di estrarre informazioni da database strutturati (*Knowledge Discovery in Database*), le tecniche di Text Mining rendono possibile l'estensione di tale approccio ad informazioni non strutturate (documenti di testo) in possesso di un'organizzazione (*Knowledge Discovery in Text*).

Nel processo di produzione di conoscenza il Text Mining rappresenta dunque un risorsa in più a disposizione del *Knowledge Information Worker*, uno strumento il cui *ruolo* può essere non solo alternativo a quello del più tradizionale data mining, ma in taluni casi complementare.

Nello specifico verrà affrontato nell'intervento un particolare esempio di elaborazione di testo: l'analisi delle domande aperte contenute in un questionario strutturato.

Nell'ambito delle ricerche di tipo survey l'utilizzo di domande aperte è limitato dalle difficoltà che si incontrano in fase di analisi. La ricodifica delle risposte ottenute è infatti un compito gravoso, sia per la mole di materiale da visionare, sia spesso per la mancanza di criteri chiari di ricodifica.

Anche se la pratica suggerisce di non ricorrere spesso a domande aperte, esistono almeno tre situazioni nelle quali tale strumento risulta essere essenziale:

- Quando si vuole indagare ulteriormente la risposta ottenuta ad una domanda chiusa.
E' il caso della domanda *perché?* posta successivamente ad una domanda chiusa nella quale, ad esempio si è chiesto il grado di accordo su di una affermazione.
- Per economizzare i tempi di intervista.
Utilizzare una domanda aperta consente di evitare di sottoporre all'intervistato lunghe serie di alternative. Non si tratta quindi solo di economizzare tempo, ma anche di aumentare la qualità del dato.
- Raccogliere risposte spontanee dell'intervistato

Da quanto detto deriva la necessità di disporre di uno strumento che consenta un trattamento *economico* dei dati testuali raccolti per mezzo di domande aperte, e che permetta di elaborare in modo congiunto domande aperte e domande chiuse che compaiono nello stesso questionario.

Esiste una famiglia di tecniche che va sotto il nome di Analisi delle Corrispondenze che si presta a coprire parzialmente questa necessità.

Un esempio classico è quello proposto sin dagli anni '80 da Lebart e Salem nell'ambito della scuola di Benzecri (L.Lebart, A. Salem Paris Dunod, *Analyse statistique des données textuelles: question ouvertes et lexicométrie*, 1998).

L'Analisi delle Corrispondenze Lessicali (ACL) consente di visualizzare l'associazione tra elementi testuali (parole o segmenti) contenuti in una particolare tabella di contingenza (tabella di Burt).

Il punto di partenza è una matrice in forma disgiuntiva completa nella quale ogni colonna rappresenta una modalità di ciascuna domanda chiusa utilizzata o una unità di testo occorrente nelle risposte a domande aperte. Le unità di testo considerate possono essere singole parole (meglio lemmi) o segmenti di testo (sequenze di parole). Le righe della matrice sono i casi della ricerca. In questo modo è possibile evidenziare per ciascun caso quali modalità di risposta a domande chiuse ha scelto e quali parole ha utilizzato nel rispondere alle domande aperte.

A partire da questa matrice l'analisi delle corrispondenze consente di studiare l'associazione tra elementi testuali e categorie di risposta rilevate per mezzo di domande chiuse, fornendo un valido aiuto nell'interpretazione delle categorie di queste ultime. Inoltre, potendo lavorare anche sulle osservazioni, è possibile derivare dalle analisi precedenti delle tipologie di intervistati (es: clusterizzazione della clientela)

Forse, più di questa breve spiegazione teorica, possono valere degli esempi per illustrare la potenzialità di tale approccio. Verranno dunque di seguito presentati tre esempi:

- Il primo illustra l'analisi congiunta di domande chiuse e di domande aperte raccolte nel corso di una indagine telefonica (CATI) sull'atteggiamento dei residenti a Torino nei confronti dei prossimi Giochi Olimpici Invernali.

L'elaborazione congiunta di questi due tipi di dati (utilizzo complementare di tecniche di Data Mining e di Text Mining) consentirà di comprendere le diverse sfumature che caratterizzano l'atteggiamento dei torinesi; informazioni preziose per chi (*decision maker*) ha la responsabilità di curare la campagna di informazione dei Giochi Olimpici Invernali 2006.

Si arriverà in questo modo a descrivere un'area di dissenso forte ed un'area di consenso ai giochi. L'analisi delle domande aperte consentirà di specificare meglio queste diverse posizioni e quindi di capire quali contenuti possono far leva sugli incerti per portarli verso una posizione di consenso.

**DATA MINING E METODOLOGIA DELLA RICERCA SOCIALE:
LA CREAZIONE DI VALORE AGGIUNTO PER L'UTENTE**

- Quindi, rimanendo in tema, si cercherà di fornire una panoramica dell'immagine che i frequentatori la Fiera del Libro hanno dato delle Alpi attraverso tre domande aperte.
- Infine verrà presentato un caso nel quale l'elaborazione di domande aperte è stata utilizzata nell'analisi della *customer satisfaction* dei clienti di una azienda.

In questo caso, si cercherà di capire quali aspetti di una certa azienda sono particolarmente apprezzati da chi dà una valutazione complessivamente positiva dei servizi ricevuti. In questo modo si tenterà di enucleare gli elementi sui quali costruire una politica di fidelizzazione della clientela.